RCA REVIEW

BOARD OF EDITORS

Chairman

R. S. Holmes
RCA Laboratories

E. I. Anderson
Home Instruments Division

A. A. Barco
RCA Laboratories

E. D. Becken
RCA Communications, Inc.

G. H. Brown
Radio Corporation of America

A. L. Conrad
RCA Service Company

E. W. Engstrom
Radio Corporation of America

A. N. Goldsmith
Honorary Vice President, RCA

J. Hillier
RCA Laboratories

E. C. Hughes
Electronic Components and Devices

E. O. Johnson
Electronic Components and Devices

E. A. Laport
Radio Corporation of America

H. W. Leverenz
RCA Laboratories

G. F. Maedel
RCA Institutes, Inc.

W. C. Morrison
Broadcast and Communications Products Division

L. S. Nergaard
RCA Laboratories

H. F. Olson
RCA Laboratories

J. A. Rajchman
RCA Laboratories

D. F. Schmit
Radio Corporation of America

L. A. Shotliff
RCA International Division

C. P. Smith
RCA Laboratories

W. M. Webster
RCA Laboratories

Secretary

C. C. Foster
RCA Laboratories

Republication and Translation

Original papers published herein may be referenced or abstracted without further authorization provided proper notation concerning authors and source is included. All rights of republication, including translation into foreign languages, are reserved by RCA Review. Requests for republication and translation privileges should be addressed to *The Manager*.

PHYSICAL LIMITATIONS ON FREQUENCY AND POWER PARAMETERS OF TRANSISTORS

By

E. O. Johnson

RCA Electronic Components and Devices,
Somerville, N. J.

Summary—A simple analysis is used to show that the ultimate performance limits of a transistor are set by the product $Ev_s/2\pi$, where E is the semiconductor's dielectric breakdown voltage and v_s is its minority-carrier saturated drift velocity. This product, which has a value of about 2×10^{11} volts/second for silicon, emphasizes the fact that a semiconductor material has a maximum capability for energizing the electric charges that process a signal. If the operating frequency of the device is high, the frequency time period is short, and only a small amount of energy can be given to a charge carrier. Consequently, the power and power amplification must be relatively low. At low frequencies the inverse is true. In other words, device physics demands an inverse relation between frequency and power parameters that is independent of the thermal dissipation arguments commonly given to explain the trade-off between these parameters. The analysis leads to an effective means for making comparisons between existing devices.

Introduction

IT SEEMS reasonable to suppose that an ultimate limit exists in the trade-off between the volt–ampere, amplification, and frequency capabilities of a transistor. This trade-off should somehow be related to material parameters. A trade-off relation linking these parameters would be useful if it could be derived in a manner that makes it independent of device design details, but yet applicable as a practical yardstick for evaluating device design refinement. J. M. Early,[1,3] and others[4-6] have considered this subject with more emphasis

[1] J. M. Early, "Maximum Rapidly-Switchable Power Density in Junction Triodes," *IRE Trans. on Electronic Devices*, Vol. ED-6, p. 322, July 1959.

[2] J. M. Early, "Speed in Semiconductor Devices," IRE National Convention, March 1962.

[3] J. M. Early, "Structure-Determined Gain-Band Product of Junction Triode Transistors," *Proc. IRE*, p. 1924, Dec. 1958.

[4] J. M. Goldey and R. M. Ryder, "Are Transistors Approaching Their Maximum Capabilities?" International Solid-State Circuits Conference, Feb. 1963.

[5] R. L. Pritchard, "Frequency Response of Grounded-Base and Grounded-Emitter Junction Transistors," AIEE Winter Meeting, Jan. 1954.

[6] L. J. Giacoletto, "Comparative High-Frequency Operation of Junction Transistors Made of Different Semiconductor Materials," *RCA Review*, Vol. 16, p. 34, March 1955.

on device design details than upon the general physical principles. One objective of the present paper is to derive general relations that show the performance limits of transistors independently of design details. Another objective is to demonstrate the use of these relationships in comparing existing devices and predicting the trend of future developments.

First, a basic voltage-frequency relation for semiconductor devices is discussed. This relationship is then extended to include space-charge constraints on device current. This combination leads directly to a relation that links maximum volt-amperes, device impedance level, and cutoff frequency. Finally, device power gain is brought into the picture through charge-control considerations. This leads to a relation that links power gain, dynamic range, and cross-modulation characteristics.

The mode of approach is to establish upper bounds on transistor performance by developing a highly idealized and simplified device model whose performance is not likely to be surpassed by that of any attainable design, no matter how optimized or cleverly conceived. This idealized performance is expressed in terms useful as a yardstick in comparing existing devices and highlighting their shortcomings. This mode of approach should complement the more usual one wherein performance improvements are extrapolated from the existing state of the art.

Voltage–Frequency Relation

The charge-carrier transit-time cutoff frequency, f_T, of a charge-control type of device[7] such as a transistor is defined by the relation $f_T = (2\pi\tau)^{-1}$, where τ is the average time for a charge carrier moving at an average velocity, v, to traverse the emitter–collector distance, L. For a given value of L, minimum value of τ occurs when v is at its maximum possible value. For a semiconductor this maximum possible value is the saturated drift velocity v_s; for holes and electrons in common semiconductor materials[8] such as silicon and germanium v_s is approximately 6×10^6 cm/sec. This limiting velocity is reached at fields of the order of 10^4 volts/cm. With the carriers moving at velocity v_s, the transit time can be reduced even further by decreasing the distance L. The lower limit on L, however, is reached when the value of V/L, where V is the applied emitter–collector voltage, becomes equal to the dielectric breakdown field, E. This is approximately 10^5 volts/cm

[7] H. Q. Johnson and H. Rosen, "Simple General Analysis of Amplifier Devices with Emitter, Control and Collector Functions," Proc. IRE, p. 407, March 1958.

[8] E. J. Ryder, "Mobility of Holes and Electrons in High Electric Fields," *Phys. Rev.*, Vol. 90, p. 766, June 1958.

in germanium and about twice this in silicon.[9] We are thus led to conclude that the best possible trade-off between the cutoff frequency, f_T, and the maximum allowable applied voltage, V_m, is given by

$$V_m f_T = \frac{E v_s}{2\pi} = \begin{cases} 2 \times 10^{11} \text{ volts/second for silicon} \\ 1 \times 10^{11} \text{ volts/second for germanium.} \end{cases} \quad (1)$$

This relationship[10] defines the upper limit on cutoff frequency, since the minimum value of V_m must have some value, say one volt, that is sufficiently greater than thermal voltage to ensure normal transistor collector action of the base–collector junction.

Practically attainable frequencies will be substantially less than the maximum possible frequencies indicated by Equation (1) because (1) the limiting velocity, v_s, will not be reached in all parts of the charge carrier path, (2) the electric field stress will not be geometrically uniform, and (3) the practical technology of small dimensions will have its own limitations. In practical circuit applications the maximum applied collector voltage is usually kept to a value below approximately one half of V_m to provide an adequate safety margin with respect to voltage transients and the instability effects caused by stray currents flowing to the base from the collector.

With respect to item (1), today's high-performance transistor designs are such that the collector depletion-layer thickness is two or three times the base thickness. Correspondingly, the velocity v_s is reached in approximately two-thirds to three-fourths of the carrier path. Thus, v_s is reduced by a factor of about 0.7. With respect to item (2), excluding semiconductor surface effects, the average electric field in the collector depletion layer ranges from one-half to about one-third of the maximum electric field, depending upon the impurity profile. Taking into account the relatively smaller field in the base region, the combined effect of items (1) and (2) is to reduce performance to a value that, at best, will be no greater than about one-fourth of that indicated by Equation (1). If the effect of the semiconductor surface on junction electric fields is taken into account, performance could be reduced by another factor of perhaps as much as two or three, particularly in higher voltage devices. However, by using the proper

[9] J. Shields, "Breakdown in Silicon on Junctions," *Jour. Electronics and Control*, Vol 6, p 130, Feb 1959.

[10] First pointed out to author by W. M. Webster. Referred to by E. O. Johnson, "Whither the Tunnel Diode," AIEE-IRE Electron Devices Conference, Oct. 1962.

geometrical contour in the region where the junction intercepts the surface of the semiconductor, this effect can be reduced or eliminated.[11]

Table I lists V_m and f_T for a variety of available transistors made from both silicon and germanium, and spanning a wide range of frequency, current, voltage, and power capability. The other parameters listed in the table are considered later in the paper. The factual data in the table was obtained from handbooks and various commercial literature. This data, it should be noted, contains varying degrees of design and manufacturing conservatism. Figure 1 compares V_m and f_T values for these transistors with Equation (1).

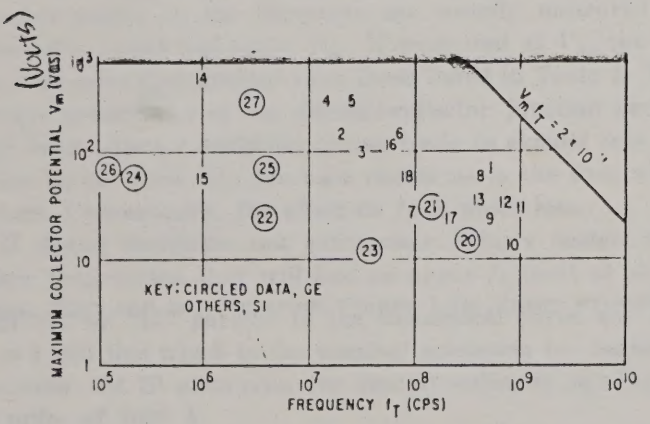


Fig. 1—Voltage–frequency relation for transistors (numbers refer to Table I).

As expected, the performance of all of the devices falls below the theoretical line. Also, as might be expected, the most recently developed devices tend to lie closest to this line. Silicon devices in almost every case are closer to the line than germanium devices. This probably stems from a lower value of E for germanium, and the fact that the germanium devices plotted in the figure are mostly of older designs. Grown-junction silicon transistors, which are the oldest devices of all and which are not shown in the figure, lie farthest away from the theoretical curve.

There is a pronounced tendency for performance to fall away from the theoretical curve in the lower frequency domain. This tendency arises from a combination of factors. Firstly, there is substantial difficulty in dealing with the lightly doped semiconductor material

[11] R. L. Davies and F. E. Gentry, "Control of Electric Field at the Surface of P-N Junctions," *IEEE Trans. on Electron Devices*, Vol. ED-11, p. 313, July 1964.

Table I—Device Data

	Device	Mat'l	Construction*	Function	Measured Data: Cutoff Frequency (cps)	Measured Data: V_m (volts)	Measured Data: I_m (amps)	Measured Data: C_c (pf)	Calculated Data: $P_m = V_m I_m$ (watts)	Calculated Data: X_c at f_T (ohms)	Calculated Data: $I_m X_c$ (volts)	Calculated Data: $(P_m X_c)^{1/2}$ (volts)	Calculated Data: $(P_m X_c)^{1/2} f_T$ (volts/sec)
1.	2N3375	Si	OE	power rf	5×10^8	65	1.5	10	97.5	32	48	56	2.8×10^{10}
2.	2N3265	Si	DDE	power rf	2×10^8	150	25	900	3750	8.85	221	182	3.6×10^9
3.	TA2400	Si	DDE	power rf	3×10^8	100	25	500	2500	11.6	290	170	5.1×10^9
4.	TA2496	Si	EM	power sw	1.5×10^8	300	10	400	3000	26.5	265	282	4.2×10^9
5.	TA2301	Si	EM	power audio	2.5×10^8	300	0.1	5	30	270	127	195	5.0×10^9
6.	TA2529	Si	EM	video driver	7×10^8	150	5×10^{-2}	2.5	7.5	910	46	82.6	5.8×10^9
7.	3N98	Si	MOS	FET	10^8	30	15×10^{-3}	2.5	0.45	1060	16	21.8	2.2×10^8
8.	2N2477	Si	DDEP	core sw	4×10^8	60	0.5	10	30.0	39.5	20	34.3	1.37×10^{10}
9.	2N2938	Si	DDEP	core sw	5×10^8	25	0.5	4	12.5	68.5	34	29.2	1.5×10^{10}
10.	2N2475	Si	DDEP	logic	8×10^8	15	0.5	3	7.5	66.5	33.3	22.4	1.8×10^{10}
11.	2N2857	Si	DDEP	rf amplifier	10^9	30	2×10^{-2}	1.8	0.6	88.5	1.8	7.3	7.3×10^9
12.	2N2708	Si	DDEP	rf amplifier	7×10^8	35	2×10^{-2}	1.5	.7	151	3	10.3	7.2×10^9
13.	TA2600	Si	OE	power rf	3.5×10^8	36	5	25	180	18.2	91	57.2	2.0×10^{10}
14.	2N689	Si	D	thyristor (SCR)	10^6	500	16	50	8000	3200	51200	5060	5.06×10^9
15.	2N1513	Si	SD	power sw	10^6	60	8	200	480	800	6400	620	6.2×10^8
16.	2N2102	Si	TDP	general	6×10^7	120	1	15	120	177	177	145	8.7×10^9
17.	2N706	Si	DDP	logic	2×10^8	25	10^{-1}	6	2.5	133	13	18.2	3.6×10^9
18.	2N696	Si	DDP	power sw	8×10^7	60	0.5	20	30	100	50	54.8	4.3×10^9
19.	2N2016	Si	D	power	2.5×10^4	130	10	400	1300	5400	1333	4500	1.1×10^8
20.	2N960	Ge	EM	logic	3×10^8	15	10^{-1}	4	1.5	1333	13	14.1	4.2×10^9
21.	2N1177	Ge	AD	rf amplifier	1.4×10^8	30	10^{-2}	2	0.3	5700	5.7	13.1	1.8×10^9
22.	2N404	Ge	A	logic	4×10^8	25	10^{-1}	12	2.5	3310	331	91	3.6×10^8
23.	2N1301	Ge	AM	logic	3.5×10^7	13	10^{-1}	8	1.3	570	57	27.2	9.5×10^8
24.	2N301	Ge	A	power	2×10^8	60	10	300	600	2650	26500	1250	2.5×10^8
25.	2N2147	Ge	DA	power	4×10^8	75	5	400	375	1000	500	194	7.8×10^8
26.	2N1358 (174)	Ge	A	power	10^8	70	15	400	1050	4000	60000	2050	2×10^8
27.	TA1928	Ge	DA	power sw	3×10^8	320	10	300	3200	1777	1800	820	2.5×10^9

* KEY:
OE—overlay epitaxial
DDE—double-diffused epitaxial
EM—epitaxial mesa
MOS—metal-oxide-semiconductor (field effect transistor)
DDEP—double-diffused epitaxial planar
D—diffused
SD—single-diffused
TDP—triple-diffused planar
DDP—double-diffused planar
AD—alloy-diffused
A—alloy
AM—alloy mesa
DA—drift alloy

necessary for high collector-breakdown voltages. This stems both from the bulk material itself, and from the relatively greater surface effects experienced with lightly doped material. Secondly, the greatest amount of engineering effort, particularly developmental, has been applied in the high-frequency, relatively low-voltage domain. Thirdly, many low-frequency devices are not specifically designed to accommodate high voltages, since such voltages are not required in most typical applications.

To some degree the locations of points on Figure 1 are pessimistic, particularly for the Ge alloy devices. The values of their cutoff frequencies quoted in the literature are usually measured at collector potentials several-fold below V_m. If measured at V_m, the values of f_T would be somewhat greater than those listed in Table I. This effect is far less pronounced in the diffused-collector junction devices because their base–collector depletion region tends to expand into the collector region rather than into the base region as is the case with the alloy devices. Consequently, the effect on f_T is much less.

If device designers can extrapolate today's design perfection to higher frequencies, they will find an upper f_T limit at about 20 gigacycles. This can be seen from Figure 1 by linear extrapolation from point "1" or "11" parallel to the theoretical curve and down to the $V_m = 1$ volt line which is the nominal minimum for normal transistor operation. At 20 gigacycles the emitter–collector spacing must be of the order of 1000 Å.

The thyristor (SCR) (point "14") was included for comparison purposes that become clear in the next section. The developmental field-effect transistor (point "7") was included because it is a charge-control device and can be directly compared with the conventional bipolar transistor.[7]

Current–Frequency Relations

The load current, I, through a charge-control device is basically defined by

$$I = \frac{Q}{\tau},$$

where Q is the total mobile charge in the emitter–collector region en route to the collector. If τ is considered as being a fixed quantity, then the ratio I/Q will also be a constant;

$$\frac{I}{Q} = \frac{1}{\tau} = \text{constant}, \tag{3}$$

and I will increase linearly with Q. In a transistor the largest possible value of I will be attained when Q is approximately equal to the total fixed charge Q_f in the emitter–collector space. If Q were somehow made to exceed Q_f, the current I would become space-charge limited, as in a vacuum tube, and the controlling action of the base would be lost. This condition sets an absolute upper limit on I.

In practice this limit is not reached, because the mobile-charge density first exceeds the fixed-charge density in local regions of the emitter–collector path, for example, at the collector edge of base. This leads, as is well known, to the frequency-reducing phenomenon of base widening. It can also lead to an increase in the electric field in the collector depletion layer, thus reducing the maximum voltage that the device will tolerate before field breakdown occurs. The maximum rated current, I_m, of a device may be well below the maximum possible current for other reasons (e.g., power dissipation, second breakdown effects, and drop-off in current gain due to decreased emitter efficiency).

If we define the maximum current as the current I_m that causes the onset of significant base widening, Equation (3) can be written in terms of V_m and a capacitance approximately equal to the usually QUOTED VALUE OF THE COLLECTOR-BASE CAPACITANCE, C_0:

$$\frac{I_m}{C_0} = \frac{V_m}{\tau_C}, \tag{4}$$

where τ_C approximates the collector depletion-layer transit time. Note that the current/capacitance ratio is ideally invariant with collector area, doping level, and parallel circuit configurations of devices. For example, an increase in the doping level in the base will give a higher value of C_0, and this will allow a proportionally larger value of I_m. Equation (4) has some interesting consequences. First of all, if we replace V_m by its value from Equation (1) we see that

$$\frac{I_m}{C_0} = \frac{Ev_s}{2\pi f_T \tau_C} = Ev_s, \tag{5}$$

since $2\pi f_T \tau_C \cong 2\pi f_T \tau = 1$. The quantity I_m/C_0, a convenient measure of a device's speed in switching voltage, has a maximum attainable value that is equal to Ev_s.

Secondly, if we convert C_0 to reactive impedance X_c at the frequency f_T by using the relation $C_0 = (2\pi f_T X_c)^{-1}$, we obtain the relation

$$(I_m X_c) f_T = \frac{E v_s}{2\pi}. \tag{6}$$

This gives the most optimistic trade-off between current and frequency that one can expect from a transistor. A comparison of the devices listed in Table I with Equation (6) is shown in Figure 2. In marked contrast with Figure 1, we note that the data does not fall away from

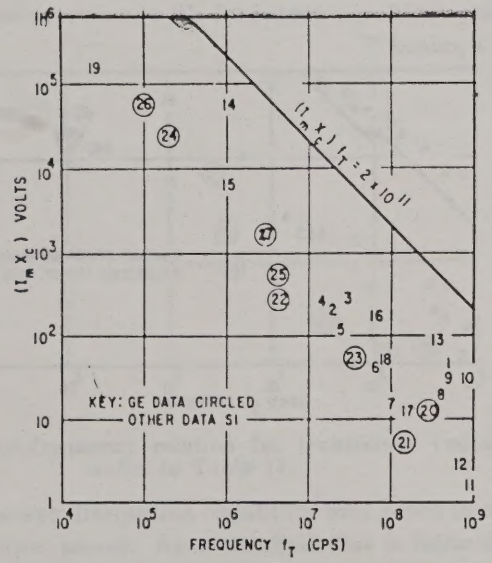


Fig. 2—Current–frequency relation for transistors (numbers refer to Table I).

the theoretical curve in the low-frequency domain. Existing devices thus appear better at handling high currents than they are at handling high voltages. Devices having an emitter geometry that maximizes the emitter periphery–area ratio (points "1", "9", "10", and "13", for example) minimize current-crowding effects and so tend to be closer to the theoretical curve.

Device current performance falls below the theoretical curve for reasons in addition to those listed above for the current itself. One reason is that because of current-crowding effects at the emitter periphery, the portions of the collector under the center of the emitter the emitter, such as those regions under base contracts, are also useless

for handling load current. For these reasons the actual collector capacitance tends to be several-fold larger than is necessary for handling the rated value of collector current. Package and other stray capacitances, particularly with the high-frequency devices, also have the same deleterious effect.

As might be expected, the thyristor (SCR) device (point "14") is closer to the theoretical curve than any of the other devices. This device is not limited to a fixed I/C ratio because the maximum collector current can be almost arbitrarily large. Collector-current space-charge effects are minimized by the counterflow of majority carriers of opposite sign. The counter-flow of carriers of opposite sign creates a dense solid-state plasma that can carry currents not directly limited by the density of fixed charge in the device. Accordingly, a thyristor is inherently superior to a transistor in its volts/second switching capability. This superiority is further enhanced by the absence of current-crowding effects; base current is supplied from the collector rather than through the lateral base resistance as in a transistor. The price one pays for the inherent switching superiority of the thyristor is that current turn-off action becomes cumbersome and relatively slow. The f_T value noted for the thyristor in Table I was calculated from the turn-on time.

Volt–Ampere–Frequency Relations

The relation between volt–ampere product, impedance level, and frequency for a transistor is obtained by multiplying Equations (1) and (6). This combination gives

$$(P_m X_c)^{1/2} f_T = \frac{E v_s}{2\pi}, \tag{7}$$

wherein the volt–ampere product, $V_m I_m$, has been replaced by P_m. J. M. Early[1,2] has shown that the left-hand side of this relation can be derived by simple intuitive arguments. One interesting conclusion from Equation (7) is that for a given device impedance, the volt–ampere ability of a device must necessarily decrease as the device cutoff frequency, f_T, is increased.*

This decrease must occur, not because of the power dissipation arguments usually cited, but because a given semiconductor material has a limited volts/second capability for the charge carriers traversing it. For a device designed to operate at a low frequency, the time period

* First pointed out to the author by J. Olmstead and J. Avins of RCA.

Accordingly, energy transfer and power capability can be relatively large. For high-frequency devices the inverse is true. More specifically, the maximum electric field to which a charge carrier in the device can be subjected is the breakdown field, E. A charge carrier can traverse this field at a maximum velocity equal to v_s. The product Ev_s is the maximum rate at which a carrier can acquire volts of energy. If time is long, then the energy transfer can be large; conversely, if the time is short, the energy transfer is restricted.

The power-dissipation argument, sometimes inferred as a principle, holds that device size decreases with frequency capability and that this

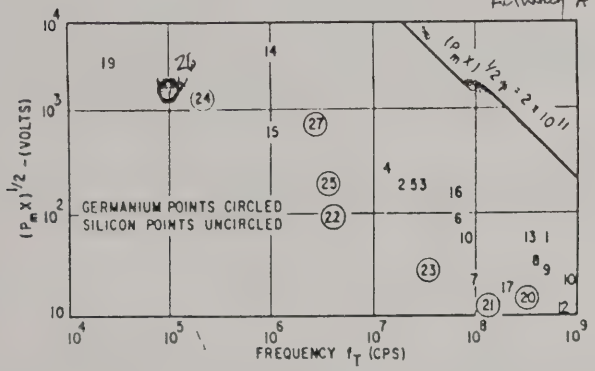


Fig. 3—Power–frequency relation for transistors (numbers refer to Table I).

leads to decreased power dissipation capability and hence to a decreased ability to handle output power. As a principle this is fallacious. Equation (7) implies that, in principle, an arbitrarily large value of P_m can be achieved, for a given f_T, by connecting devices in parallel. The price paid for this is a proportionate decrease in the impedance X_c. In principle, these devices can be physically arranged in a distributed array such that the heat dissipation capability can be made almost arbitrarily high. Indeed, future practical power-output limits may well be established more by the problems of economics, circuitry, and device uniformity than by heat dissipation, per se.

Equation (7) points out that when the best power–frequency capability is desired for a transistor, the transistor should be designed to operate at relatively high current. This choice is a trade for a decrease in impedance level, but not for a decreased frequency capability as would be the case if the device were designed for low current and high voltage. Circuit requirements and practical device design problems limit how far one can proceed in the direction of high current.

Figure 3 compares the performance of the devices in Table I with

Equation (7). Again, as in Figure 1 and for the same reasons, performance tends to diverge more from the theoretical at the lower f_T values. Also, the superiority of silicon devices compared to germanium devices is strongly implied. The most recently developed designs for silicon power transistors, which use a large emitter periphery–area ratio, are closest to the theoretical curve for the reasons given previously.

Relations Involving Amplification Properties

The maximum available power gain, G_p, of a charge-control device such as a transistor, is given by[7]

$$G_p = G_I^2 \frac{Z_0}{Z_{in}} = G_I^2 \frac{Z_0}{Z_{IN}} = G_P = \left(\frac{f_T}{f}\right)^2 \frac{Z_0}{Z_{in}}, \tag{8}$$

where G_I is the current-amplification ($G_I = (2\pi f\tau)^{-1} = f_T/f$), and Z_0 and Z_{in} are the output and input impedances, respectively. As noted here, G_p refers to the power gain that would be measured over the low-frequency pass band when the load is resistive. If we make the idealizing assumption that no electrode series resistances exist, Equation (8) can be approximated by

$$G_p = \left(\frac{f_T}{f}\right)^2 \frac{C_{in}}{C_0}, \tag{9}$$

where C_{in} is the input capacitance and C_0 is the output (base–collector) capacitance. Furthermore, if we assume that the emitter diffusion capacitance, C_d, dwarfs the emitter transition capacitance, then the maximum value of the input capacitance is given by

$$C_{in} = C_d \cong \frac{Q_m}{V_T} = \frac{I_m \tau_b}{V_T}, \tag{10}$$

where Q_m is the maximum total carrier charge en route to the collector, τ_b in the carrier base transit time, and V_T is the thermal voltage (kT/e volts). Other quantities have been previously defined. The output capacitance, C_0, is found from Equation (4):

$$C_0 = \frac{I_m \tau_c}{V_m}. \tag{11}$$

The maximum capacitance ratio is thus defined by Equations (10) and (11), and Equation (9) can be rewritten

$$G_p = \left(\frac{f_T}{f}\right)^2 \frac{V_m}{V_T} \frac{\tau_b}{\tau_c} \simeq \left(\frac{f_T}{f}\right)^2 \frac{V_m}{V_p}. \tag{12}$$

The frequency ratio accounts for the input–output charge amplification of the device, and the voltage ratio, for the energy step-up per carrier. That is, in the ideal case an energy of V_T is required to place a charge on the control electrode, the base. As a consequence, each of the charges that flows to the collector ideally picks up an energy equal to V_m.

Equation (12) can be written in a form consistent with the other performance trade-off relations by using Equation (1);

$$(G_p V_T V_m)^{1/2} f = \frac{E v_s}{2\pi}. \tag{13}$$

Here G_p is evaluated at some operating frequency, f, sufficiently close to f_T for Equation (8) to be valid. The value of V_m corresponds to that used in Equation (1). Equation (13) emphasizes the fact that power amplification, like volt–ampere performance, basically depends upon the volts/second capability of the semiconductor material.

The actual performance of a transistor will fall far short of that predicted by Equation (13). First of all, the qualifications on practically attainable electric fields and average carrier velocities noted for Equation (1) will apply. This reduces the right-hand constant to an effective value equal to about one fourth, or less, of the theoretical value. Secondly, the practical working value of applied collector voltage will be about one half of V_m (see Table II). Thirdly, Equation (10) overestimates the value of C_d and, hence, G_p by perhaps as much as two times in a device operating at very high carrier densities. This disparity stems from an electric-field-enhanced carrier velocity in the base. Fourthly, this same electric field in the base doubles the input-energy requirement per carrier. Fifthly, another factor of at least two is added because of ohmic base resistance. The energy lost in the resistance of a series RC circuit is equal to the potential energy acquired by the charges on the capacitor when the charging time is more than a few-fold larger than the RC time constant. The energy loss in the base resistance can be very much greater than this when the driving signal's time constant is less the RC time constant of the circuit. Transition capacitance accentuates the problem because it increases the time constant of the transistor's input circuit. The effect of all these factors is to make the right-hand constant in Equation (13) at least thirty-fold less than the theoretical constant.

Table II

Device	Material	Construction*	Measured Data: Operating Freq. (f) (cps)	Power Gain (G_p) (db)	Power Gain (G_p) (ratio)	Voltage (volts) Operating	Voltage (volts) Max. (V_m)	Calculated Data: $(V_m V_T G_p)^{1/2}$ (volts)
1. 2N1631	Ge	DA	1.5×10^6	47.7	6.0×10^4	12	34	232
2. 2N1425	Ge	DA	4.6×10^5	51.0	1.3×10^5	12	24	280
3. 2N1180	Ge	DA	1.0×10^7	35.0	3.2×10^3	12	30	15.5
4. 2N1638	Ge	DA	2.6×10^5	61.5	1.4×10^6	11	34	1100
5. 2N384	Ge	DA	5.0×10^7	21.0	1.3×10^2	12	40	11.4
6. 2N1066	Ge	DA	5.0×10^7	26.0	4.0×10^2	12	40	20
7. 2N1177	Ge	DA	1.0×10^8	14.0	2.5×10^1	12	30	4.3
8. 2N175	Ge	A	2.0×10^4	43.0	2.0×10^4	4	10	71
9. 2N139	Ge	A	4.6×10^5	37.0	5.0×10^3	9	16	45
10. 2N408	Ge	A	4.6×10^5	37.8	6.0×10^3	9	20	55
11. 2N2873	Ge	M	1.8×10^8	21.6	1.5×10^2	12	35	11.4
12. 2N2708	Si	DDEP	2.0×10^8	22.0	1.6×10^2	15	35	11.8
13. 2N2875	Si	DDEP	4.5×10^8	19.0	8.0×10^1	6	30	7.8
14. 3N98	Si	MOS	6.0×10^7	10.0	10	20	30	34.6

*KEY:
DA—drift alloy
A—alloy
M—mesa
DDEP—double-diffused epitaxial planar
MOS—*Metal Oxide Semiconductor*—field effect transistor
As noted in text, the voltage corresponding to V_T for the MOS is approximately the pinch-off voltage. Value used in calculation was 4.0 volts.

A plot of data in Table II is shown in Figure 4 along with the theoretical curve corresponding to Equation (13). The points closest to the theoretical curve correspond to the most recently developed devices and are roughly a hundred-fold removed. The next closest group of points, "1" through "7", correspond to the next most recently developed devices, the alloy drift types. The farthest removed group of points, "8", "9", and "10", correspond to the oldest device types, the alloy devices.

One reason for the tendency for the points to diverge increasingly from the theoretical curve at lower frequencies has to do with the previously discussed factors involving V_m, which apply to the results

in Figure 1. Another reason is that circuit instability problems will tend to limit the amount of power gain sought by the device designer.

Making a speculative extrapolation to higher frequencies with Figure 4, as was done for the data in Figure 1, we see that at an operating frequency of 20 gigacycles the value of $(G_p V_T V_m)^{1/2}$ will be somewhat less than unity, and G_p will thus be limited to a maximum value of about ten.

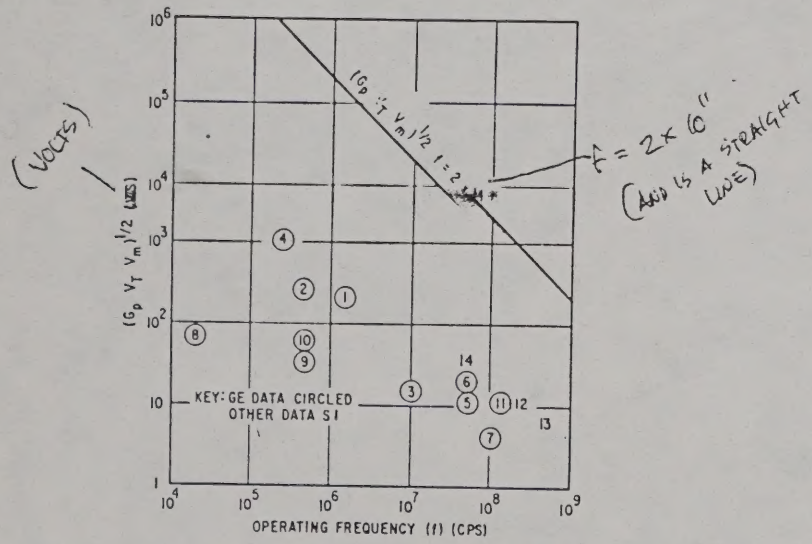


Fig. 4—Power-gain and frequency relation for transistors (numbers refer to Table II).

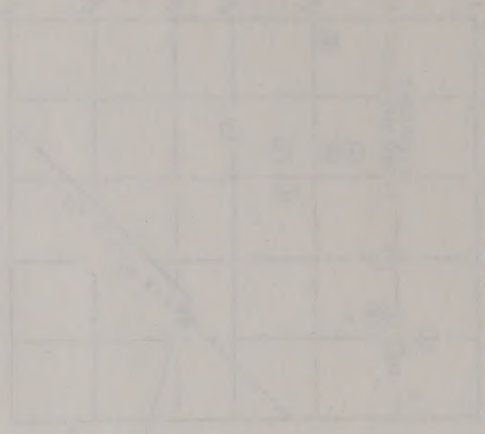
Equations (12) and (13) suggest that one must trade power gain for the energy required to place a charge on the control electrode.[7] For the bipolar transistor, this input energy has been noted as V_T volts; for the field-effect transistor, this energy can be associated with the pinch-off voltage, V_p, or the G_m/I ratio, and is ten to a hundred times greater than V_T. Both V_p and V_T define approximately the useful dynamic input voltage range of field effect and bipolar devices. Simple consideration shows that the larger this dynamic input-voltage range, the smaller the curvature components on the transfer characteristics tend to be. Consequently, the tendency for cross-modulation phenomena in the device will be smaller, particularly when the device operating point is swept along the transfer characteristic during reverse a-g-c action.*

* Varactor action in the collector circuit is not included in this argument.

One concludes that a bipolar transistor will tend to have more power gain at a given operating frequency than a field-effect transistor when both devices are constructed to have the same value of f_T. On the other hand, the field-effect transistor will have a greater dynamic range and, correspondingly, a greater ability to avoid cross-modulation interference effects.[12] Point "14" in Figure 4 shows that the field-effect transistor holds its own against the bipolar transistor when the dynamic range factor is taken into account.

Conclusions

A simple analysis of transistors shows that the product of the semiconductor breakdown field and the maximum carrier drift velocity, Ev_s, is the ultimate measure of a transistor's volt–ampere, power gain, and frequency performance. This product emphasizes that a given semiconductor material used in a transistor has a definite and fixed capability for imparting volts of energy per second to a charge carrier. For this reason the energy transfer to a charge carrier must necessarily decrease with frequency as must the volt–ampere and power-gain performance. However, if the designer is willing to pay the price of a decreased impedance level, a transistor or transistor array can, in principle, be made to give an almost arbitrarily high volt–ampere performance at any operating frequency within the frequency limits of the device.

For a given device cutoff frequency, f_T, there is a trade-off between power gain and dynamic range. The gain of a bipolar transistor will tend to be superior to that of the field-effect transistor, while the field-effect transistor will tend to have a superior dynamic range and ability to avoid cross-modulation interference.

The various trade-off relations can be plotted in a simple manner that compares transistors having a very wide range of parameter values.